

Evaluating the Impact of Color on Texture Recognition

Fahad Shahbaz Khan¹, Joost van de Weijer², Sadiq Ali³, and Michael Felsberg¹

¹ Computer Vision Laboratory, Linköping University, Sweden

fahad.khan@liu.se

² Computer Vision Center, CS Dept. Universitat Autònoma de Barcelona, Spain

³ SPCOMNAV, Universitat Autònoma de Barcelona, Spain

Abstract. State-of-the-art texture descriptors typically operate on grey scale images while ignoring color information. A common way to obtain a joint color-texture representation is to combine the two visual cues at the pixel level. However, such an approach provides sub-optimal results for texture categorisation task.

In this paper we investigate how to optimally exploit color information for texture recognition. We evaluate a variety of color descriptors, popular in image classification, for texture categorisation. In addition we analyze different fusion approaches to combine color and texture cues. Experiments are conducted on the challenging scenes and 10 class texture datasets. Our experiments clearly suggest that in all cases color names provide the best performance. Late fusion is the best strategy to combine color and texture. By selecting the best color descriptor with optimal fusion strategy provides a gain of 5% to 8% compared to texture alone on scenes and texture datasets.

Keywords: Color, texture, image representation.

1 Introduction

Texture categorisation is a difficult task. The problem involves assigning a class label to the texture category it belongs to. Significant amount of variations in images of the same class, illumination changes, scale and viewpoint variations are some of the key factors that make the problem challenging. The task consists of two parts, namely, efficient feature extraction and classification. In this work we focus on obtaining compact color-texture features to represent an image.

State-of-the-art texture descriptors operate on grey level images. Color and texture are two of the most important low level visual cues for visual recognition. A straight forward way to extend these descriptors with color is to operate on separately on the color channels and then concatenate the descriptors. However such representations are high dimensional. Recently, it has been shown that an explicit color representation improves performance on object recognition and detection tasks [1,2]. Therefore, this work explores several pure color descriptors popular in image classification for texture categorisation task.

There exist two main approaches to combine color and texture cues for texture categorisation.

Early Fusion: Early fusion fuses the two cues at the pixel level to obtain a joint color-texture representation. The fusion is obtained by computing the texture descriptor on the color channels. Early fusion performs best for categories which exhibit constancy in both color and shape [1].

Late Fusion: Late fusion process the two visual cues separately. The two histograms are concatenated into a single representation which is then the input to a classifier. Late fusion combines the visual cues at the image level. Late fusion works better for categories where one cue remains constant and the other changes significantly [1]. In this work we analyze both early and late fusion approaches for the task of texture categorisation.

As mentioned above, state-of-the-art early fusion approaches [3] combine the features at the pixel level. Contrary to computer vision, it is well known that visual features are processed separately before combining at a later stage for visual recognition in human brain [4,5]. Recently, Khan et al. [6] propose an alternative approach to perform early fusion for object recognition. The visual cues are combined in a single product vocabulary. A clustering algorithm based on information theory is then applied to obtain a discriminative compact representation. Here we apply this approach to obtain a compact early fusion based color-texture feature representation.

In conclusion, we make the following novel contributions:

- We investigate state-of-the-art color features used for image classification for the task of texture categorisation. We show that the color names descriptor with its only 11 dimensional feature vector provides the best results for texture categorisation.
- We analyze fusion approaches to combine color and texture. Both early and late feature fusion is investigated in our work.
- We also introduce a new dataset of 10 different and challenging texture categories as shown in Figure 1 for the problem of color-texture categorisation. The images are collected from the internet and Corel collections.

2 Relation to Prior Work

Image representations based on color and texture description are an interesting research problem. Significant amount of research has been done in recent years to the solve the problem of texture description [7,8,9,10]. Texture description based on local binary patterns [8] is one of the most commonly used approach for texture classification. Other than texture classification, local binary patterns have been employed for many other vision tasks such as face recognition, object and pedestrian detection. Due to its success and wide applicability, we also use local binary patterns for texture categorisation in this paper¹.

¹ We also investigated other texture descriptors such as MR8 and Gabor filters but inferior results were obtained compared to LBP. However, the approach presented in this paper can be applied with any texture descriptor.

Color has shown to provide excellent results for bag-of-words based object recognition [3,1]. Recently, Khan et al. [1,2] have shown that an explicit representation based on color names outperforms other color descriptors for object recognition and detection. However, the performance of color descriptors, popular in image classification, has yet to be investigated for texture categorization task. Therefore, in this paper we investigate the contribution of color for texture categorization. Different from the previous methods [11,12], we propose to use color names as a compact explicit color representation. We investigate both late and early fusion based global color-texture description approaches. Contrary to conventional pixel based early fusion methods, we use an alternative approach to construct a compact color-texture image representation.

3 Pure Color Descriptors

Here we show a comparison of pure color descriptors popular in image classification for texture description.

RGB Histogram [3]: As a baseline, we use the standard RGB descriptor. The RGB histogram combines the three histograms from the R, G and B channels. The descriptor has 45 dimensions.

rg Histogram [3]: The histogram is based on the normalized RGB color model. The descriptor is 45 dimensional and invariant to light intensity changes and shadows.

C Histogram: This descriptor has shown to provide excellent results on the object recognition task [3]. The descriptor is derived from the opponent color space as $\frac{O_1}{O_3}$ and $\frac{O_2}{O_3}$. The channels O_1 and O_2 describe the color information. Whereas O_3 channel contains the intensity information in an image. We quantize the descriptor into 36 bins using K-means to construct a histogram.

Opponent-angle Histogram [13]: The opponent-angle histogram proposed by van de Weijer and Schmid is based on image derivatives. The histogram has 36 dimensions.

HUE Histogram [13]: The descriptor was proposed by [13] where hue is weighted by the saturation of a pixel in order to counter the instabilities in hue. This descriptor also has 36 dimensions.

Transformed Color Distribution [3]: The descriptor is derived by normalizing each channel of RGB histogram. The descriptor has 45 dimensions and is invariant to scale with respect to light intensity.

Color Moments and Invariants [3]: In the work of [3] the color moment descriptor is obtained by using all generalized color moments up to the second degree and the first order. Whereas color moment invariants are constructed using generalized color moments.

Hue-saturation Descriptor: The hue-saturation histogram is invariant to luminance variations. It has 36 dimensions (nine bins for hue times four for saturation).

Color Names [14]: Most of the aforementioned color descriptors are designed to achieve photometric invariance. Instead, color names descriptor balances a certain degree of photometric invariance with discriminative power. Humans use color names to communicate color, such as “black”, “blue” and “orange”. In this work we use the color names mapping learned from the Google images [14].

4 Combining Color and Texture

Here we discuss different fusion approaches to combine color and texture features.

Early Fusion: Early fusion involves binding the visual cues at the pixel level. A common way to construct an early fusion representation is to compute the texture descriptor on the color channels. Early fusion results in a more discriminative representation since both color and shape are combined together at the pixel level. However, the final representation is high dimensional. Constructing an early fusion representation using color channels with a texture descriptor for an image I is obtained as:

$$T_E = [T_R, T_G, T_B], \quad (1)$$

Where T can be any texture descriptor. Most color-texture approaches in literature are based on early fusion approach [12,3]. Recently, Khan et al. [1] have shown that early fusion performs better for categories that exhibit constancy of both color and shape. For example, the foliage category has a constant shape and color.

Late Fusion: Late fusion involves combining visual cues at the image level. The visual cues are processed independently. The two histograms are then concatenated into a single representation before the classification stage. Since the visual cues are combined at the histogram level, the binding between the visual cues is lost. A late fusion histogram for an image is obtained as,

$$T_L = [H_T, H_C], \quad (2)$$

Where H_T and H_C are explicit texture and color histograms. Late fusion provides superior performance for categories where one of the visual cues changes significantly. For example, most of the man made categories such as car, motor-bike etc. changes significantly in color. Since an explicit color representation is used for late fusion, it is shown to provide superior results for such classes [1].

Portmanteau Fusion: Most theories from the human vision literature suggest that the visual cues are processed separately [4,5] and combined at a later stage for visual recognition. Recently, Khan et al. [6] propose an alternative solution for constructing compact early fusion within the bag-of-words framework. Color and shape are processed separately and a product vocabulary is constructed. A Divisive information theoretic clustering algorithm (DITC) [15] is then applied to obtain a compact discriminative color-shape vocabulary. Similarly, in this

work we also aim at constructing a compact early fusion based color-texture representation².

Here we construct separate histograms for both color and texture and product histogram is constructed. Suppose that $T = \{t_1, t_2, \dots, t_L\}$ and $C = \{c_1, c_2, \dots, c_M\}$ represent the visual texture and color histograms, respectively. Then the product histogram is given by

$$TC = \{tc_1, tc_2, \dots, tc_S\} = \{\{t_i, c_j\} \mid 1 \leq i \leq L, 1 \leq j \leq M\},$$

where $S = L \times M$. The product histogram is equal to number of texture bins times number of color histogram bins. This leads to high dimensional feature representation. This product histogram is then input to the DITC algorithm to obtain a low dimensional compact color-texture representation. The DITC algorithm works on the class-conditional distributions over product histograms. The class-conditional estimation is measured by the probability distribution $p(R|tc_s)$, where $R = \{r_1, r_2, \dots, r_O\}$ is the set of O classes. The DITC algorithm works by estimating the drop in mutual information I between the histogram TC and the class labels R . The transformation from the original histogram TC to the new representation $TC^R = \{TC_1, TC_2, \dots, TC_J\}$ (where every TC_j represents a group of clusters from TC) is equal to

$$\begin{aligned} & I(R; TC) - I(R; TC^R) \\ &= \sum_{j=1}^J \sum_{tc_s \in TC_j} p(tc_s) KL(p(R|tc_s), p(R|TC_j)), \end{aligned} \quad (3)$$

where KL is the Kullback-Leibler divergence between the two distributions defined by

$$KL(p_1, p_2) = \sum_{y \in Y} p_1(y) \log \frac{p_1(y)}{p_2(y)}. \quad (4)$$

The algorithm finds a desired number of histogram bins based on minimizing the loss in mutual information between the bins of product histogram and the class labels of training instances. Histogram bins with similar discriminative power are merged together over the classes. We refer to Dhillon et al. [15] for a detail introduction on the DITC algorithm.

5 Experimental Results

To evaluate the performance of our approach we have collected a new dataset of 400 images for color-texture recognition. The dataset consists of 10 different

² In our experiments we also evaluated PCA and PLS but inferior results were obtained. A comparison of other compression techniques with DITC is also performed by [16].

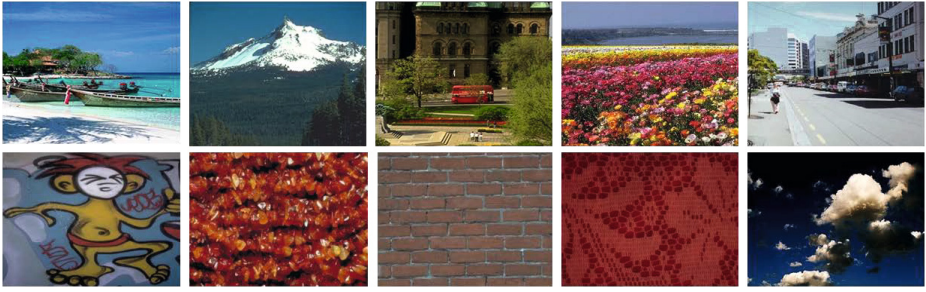


Fig. 1. Example images from the two datasets used in our experiments. First row: images from the OT scenes dataset. Bottom row: images from our texture dataset.

categories namely: marble, beads, foliage, wood, lace, fruit, cloud, graffiti, brick and water. We use 25 images per class for training and 15 instances for testing. Existing datasets are either grey scale, such as the Brodatz set, or too simple, such as the Outex dataset, for color-texture recognition. Texture cues are also used frequently within the context of object and scene categorisation. Therefore, we also perform experiments on the challenging OT scenes dataset [17]. The OT dataset [17] consists of 2688 images classified as 8 categories. Figure 1 shows example images from the two datasets.

In all experiments a global histogram is constructed for the whole image. We use LBP with uniform patterns having final dimensionality of 383. Early fusion is performed by computing the texture descriptor on the color channels. For late fusion, histograms of pure color descriptor is concatenated with a texture histogram. A non-linear SVM is used for classification. The performance is evaluated as a classification accuracy which is the number of correctly classified instances of each category. The final performance is the mean accuracy obtained from all the categories. We also compare our approach with color-texture descriptors proposed in literature [12,10].

Table 1. Classification accuracy on the two datasets. (a) Results using different pure color descriptors. Note that on both datasets color names being additionally compact provides the best results. (b) Scores using late fusion approaches. On both datasets late fusion using color names provides the best results while being low dimensional.

Method	Size	OT [17]	Texture	Method	Size	OT [17]	Texture
RGB	45	43	51	RGB LBP	383 + 45	79	74
rg	30	39	50	rg LBP	383 + 30	80	69
HUE	36	38	43	HUE LBP	383 + 36	80	74
C	36	39	41	C LBP	383 + 36	79	73
Opp-angle	36	33	27	Opp-angle LBP	383 + 36	79	74
Transformed color	45	40	41	Transformed color LBP	383 + 45	79	72
Color moments	30	42	50	Color moments LBP	383 + 30	80	74
Color moments inv	24	23	34	Color moments inv LBP	383 + 24	23	71
HS	36	37	42	HS LBP	383 + 36	79	72
Color names	11	46	56	Color names LBP	383 + 11	82	77

(a)

(b)

5.1 Experiment 1: Pure Color Descriptors

We start by providing results on the pure color descriptors discussed in Section 3. The results are presented in Table 1. On both datasets, the baseline RGB provides improved results compared to several other sophisticated color descriptors. Among all the descriptors, the color names descriptor provides best results on both datasets. Note that color names being additionally compact, possesses a certain degree of photometric invariance together with discriminative power. It has the ability to encode achromatic colors such as grey, white etc. Based on these results, we propose to use color names as an explicit color representation to combine with texture cue.

5.2 Experiment 2: Fusing Color and Texture

Here, we first show results obtain by late fusion approaches in Table 1. The texture descriptor with 383 dimensions provides a classification score of 77% and 69% respectively. The late fusion of RGB and LBP provides a classification score of 79% and 74%. The STD [12] descriptor provides inferior results of 58% and 67% respectively. The best results are obtained on both datasets using the combination of color names with LBP. Table 2 shows results obtained using early fusion approaches on the two datasets. The conventional pixel based descriptors provide inferior results on both datasets. The LCVBP descriptor [10] provides classification scores of 76% and 53% on the two datasets. By taking the product histogram directly without compression provides an accuracy of 81% and 72% while being significantly high dimensional. It is worthy to mention that both JTD and LCVBP descriptors are also significantly high dimensional. The portmanteau fusion provides the best results among early fusion based methods while additionally being compact in size.

In summary late fusion provides superior performance while being compact on both datasets. Among early fusion based methods portmanteau fusion provides improved performance on both datasets. The best results are achieved using the color names descriptor. Color names having only an 11 dimensional histogram is

Table 2. Classification accuracy using early fusion approaches. Among early fusion approaches, portmanteau fusion provides the best results on both datasets while additionally being compact.

Method	Dimension	OT [17]	Texture
RGBLBP	1149	79	70
CLBP	1149	78	69
OPPLBP	1149	80	70
HSVLP	1149	78	71
JTD [12]	15625	57	61
LCVBP [10]	15104	76	53
Product	4213	81	72
Portmanteau fusion	500	82	73

compact, possesses a certain degree of photometric invariance while maintaining discriminative power. Note that in this paper we investigate global color-texture representation. Such a representation can further be combined with local bag-of-words based descriptors for further improvement in performance.

6 Conclusions

We evaluate a variety of color descriptors and fusion approaches popular in image classification for texture recognition. Our results suggest that color names provides the best performance for texture recognition. Late fusion is an optimal approach to combine the two cues. Portmanteau fusion provides superior results compared to conventional pixel level early fusion. On scenes and texture datasets, color names in a late fusion settings significantly improve the performance by 5% to 8% compared to texture alone.

Acknowledgments. We acknowledge the support of Collaborative Unmanned Aerial Systems (within the Linnaeus environment CADICS), ELLIIT, the Strategic Area for ICT research, funded by the Swedish Government, and Spanish project TIN2009-14173.

References

1. Khan, F.S., van de Weijer, J., Vanrell, M.: Modulating shape features by color attention for object recognition. *IJCV* 98(1), 49–64 (2012)
2. Khan, F.S., Anwer, R.M., van de Weijer, J., Bagdanov, A.D., Vanrell, M., Lopez, A.M.: Color attributes for object detection. In: *CVPR* (2012)
3. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. *PAMI* 32(9), 1582–1596 (2010)
4. Treisman, A., Gelade, G.: A feature integration theory of attention. *Cogn. Psych.* 12, 97–136 (1980)
5. Wolfe, J.M.: Watching single cells pay attention. *Science* 308, 503–504 (2005)
6. Khan, F.S., van de Weijer, J., Bagdanov, A.D., Vanrell, M.: Portmanteau vocabularies for multi-cue image representations. In: *NIPS* (2011)
7. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. *PAMI* 27(8), 1265–1278 (2005)
8. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI* 24(7), 971–987 (2002)
9. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. *IJCV* 62(2), 61–81 (2005)
10. Lee, S.H., Choi, J.Y., Ro, Y.M., Plataniotis, K.: Local color vector binary patterns from multichannel face images for face recognition. *TIP* 21(4), 2347–2353 (2012)
11. Topi Maenpaa, M.P.: Classification with color and texture: jointly or separately? *PR* 37(8), 1629–1640 (2004)
12. Susana Alvarez, M.V.: Texton theory revisited: A bag-of-words approach to combine textons. *PR* 45(12), 4312–4325 (2012)

13. van de Weijer, J., Schmid, C.: Coloring local feature extraction. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 334–348. Springer, Heidelberg (2006)
14. van de Weijer, J., Schmid, C., Verbeek, J.J., Larlus, D.: Learning color names for real-world applications. TIP 18(7), 1512–1524 (2009)
15. Dhillon, I., Mallela, S., Kumar, R.: A divisive information-theoretic feature clustering algorithm for text classification. JMLR 3, 1265–1287 (2003)
16. Elfiky, N., Khan, F.S., van de Weijer, J., Gonzalez, J.: Discriminative compact pyramids for object and scene recognition. PR 45(4), 1627–1636 (2012)
17. Oliva, A., Torralba, A.B.: Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV 42(3), 145–175 (2001)